

# Multi-format Notifications for Multi-tasking

Julie S. Weber<sup>1</sup>, Mark W. Newman<sup>2</sup>, and Martha E. Pollack<sup>1,2</sup>

<sup>1</sup> Computer Science & Engineering

<sup>2</sup> School of Information, University of Michigan

Ann Arbor, MI USA

{weberjs, mwnewman, pollackm}@umich.edu

**Abstract.** We studied people's perception of and response to a set of visual and auditory notifications issued in a multi-task environment. Primary findings show that participants' reactive preference ratings of notifications delivered in various contexts during experimentation appear to contradict their reflective, overall ratings of the notification formats when elicited independently of contextual information, indicating a potential difficulty in people's abilities to articulate their preferences in the absence of context. We also found people to vary considerably in their preferences for different notification formats delivered in different contexts, such that simple approaches to selecting notification delivery formats will be dissatisfying to users a substantial portion of the time. These findings can inform the designs of future systems: rather than target the general user alone, they should strive to better understand each user individually.

**Keywords:** Notification interfaces, multi-format notification, user preferences.

## 1 Introduction

With technology becoming ubiquitous and our attention constantly shifting among numerous devices and applications, notification technology is growing increasingly important. Notifications may be spawned by an instant messaging service, a calendar system, or myriad other applications that deliver new or updated information to their users. While useful, notifications also have the potential to adversely affect task performance, increase users' stress or frustration levels, and degrade the overall user experience.

Much research has addressed both the *timing* and *format* of notifications as a potential fix to these negative effects of computerized notifications, especially those that interrupt an ongoing task (e.g., [1,2,7]). Such research has focused primarily on the effects of notifications on user *performance* and only occasionally considers user *preferences* for notifications. Furthermore, when user preferences *are* included in an evaluation, they are often used as a secondary measure of a system's effectiveness, after performance.

Because it has been shown that annoyance is an underlying cause for displeasure with an intelligent computerized assistant [10], which may be grounds for discontinued use of such systems, we are interested in accounting for user preferences for notifications. We will describe two methods we use for measuring user preferences for

notifications: *reactive* ratings (positive or negative feedback issued directly upon receipt of a notification) and *reflective* ratings (a ranked list of preferred notifications, provided during post-session interviews).

Our primary hypotheses were that (1) certain contextual features, including specific features related to the tasks at hand and the notification being delivered, will influence a user's preferences for different notification formats, and (2) people will have the ability to accurately report their own notification preferences. In keeping with prior work, we regard as "contextual features" those aspects of notification delivery and a user's tasks that have been found, in various representations and combinations, to influence certain effects of a notification. These features are the cognitive task load, or attentional focus requirements, imposed by the task being performed when a notification is received, and the importance and urgency (jointly referred to as *utility*) of the topic of the notification. We ask whether these features themselves also influence the desirability of several different types of notification delivery formats: (1) occluding, or highly intrusive visual notifications; (2) peripheral, or mildly intrusive visual notifications; and (3) an auditory notification that is presented repeatedly until acknowledged.

To begin to understand the space of notification preferences as they apply to multi-format notifications delivered in a multi-task environment, we conducted a pair of experimental user studies. A preliminary study focused on a specific type of notification: reminders. It asked participants to indicate their format preferences for the delivery of reminders, both in the absence of contextual information and then via a set of hypothetical scenarios. Our second study asked users to rate three different notification delivery formats while participating in an actual multi-task setting.

Our central results indicate that contextual setting does influence the desirability of notification delivery formats, but not exclusively: people's preferences for individual notification formats vary among individuals and across contextual settings. We also found conflicting responses to the different notification formats delivered in a multi-task setting: participants' *reflective* feedback indicated an overall distaste for occluding notifications, yet that particular format was most positively rated when *reactive* feedback was provided. This conflict indicates that preferences for notification interfaces, at least in a multi-task setting, are difficult for users to articulate without access to an explicit contextual description.

Our contribution, then, is the result of a shift in focus toward the examination of *individual* user preferences, focusing on the individual rather than the general user, and studying users' affective reactions to notifications rather than any impacts on performance.

## 2 Related Work

Notifications have been evaluated both in the home and in the workplace. Vastenburg, *et al.* [15] showed that in the home, the desirability of notifications hinged on the contextual feature of urgency alone. Other researchers focus on other features, aside from urgency. In particular, McCrickard and Chewar [12] considered the potential effects of adapting notification delivery formats to contextual setting and inferred user goals. They suggest different types of animations that would be most suitable to different situations.

However, animations tend to be disruptive, as do notifications that arrive at inopportune times. Many researchers have addressed the issue of notification *timing* and how best to adjust it relative to a user's current task, e.g., [1,3]. Iqbal and Bailey [9] first developed a system to detect breakpoints within ongoing tasks, and then they examined the effects of deferring notifications to appropriate breakpoints based on the content of the notification. They found that people reacted faster and with less frustration when notifications were issued at breakpoints than at other points within the task. Fogarty, et al. [5] that simple sensors can provide accurate estimates of human interruptibility such that notifications can be scheduled accordingly. However, there may be certain situations in which the timing of a notification cannot be modified, in which case modifying notification format may be particularly useful.

Two recent studies are similar to our own in their evaluation of multi-format notifications. Gluck, *et al.* [7] studied the effects of correlating the utility of a notification with its relative level of attentional draw. While they also considered multi-format notifications, their study involved one single task per session: not a multi-task setting as investigated in our study. Consequently, they assumed that utility and attentional draw are perceived by users solely based on relevance to a user's *current task*, while we consider incoming notifications to be associated with a particular *secondary task* with the potential to be addressed concurrently. We therefore consider utility to comprise both a notification's level of importance and also the urgency with which the information within the notification must be digested.

Mark, *et al.* [11] studied the effects of different types of interruptions on user performance on a set of office tasks. They, like many of the others described above, primarily considered the disruption costs (here relative to performance and emotional effects, i.e., stress and frustration) of notifications rather than general user desirability of, or preferences for, the format of interruption delivery.

We are thus breaking new ground in studying attention management, with respect to multi-format notifications, with a primary focus on *preferences* rather than performance.

### 3 User Study 1: Situated Reminders

We performed a preliminary study, summarized here and described in more detail in [16], that laid a foundation for the experiment that is the primary focus of this paper. Sparked by informal interviews with users of intelligent systems, and public sentiment about commercial systems (e.g., [10]) this preliminary study first aimed to characterize notification delivery formats with respect to annoyance. Then, a second phase of the study was conducted to evaluate the correlation between annoyance, or intrusiveness, of a delivery format and its contextual acceptability corresponding to a given set of hypothetical scenarios.

We acquired user preference data from a set of twenty participants. In the first phase of Study 1, participants were asked to rate each of eight reminder delivery formats on their perceived annoyance level on a scale from 0 to 10. In the second phase of the study, participants were asked to classify each reminder format as being best, acceptable or unacceptable based on its content and the given situation. From this initial phase of the study, we learned that our reminder delivery formats can be

grouped into two distinct categories based on their relative levels of perceived *intrusiveness*. This categorization influenced the design decisions for our second study, described in the following section. In particular, we note that the set of “highly intrusive” reminders, or more generally notifications, included any notification delivered via an occluding window, and “moderately intrusive” notifications all appeared in the periphery of the screen.

The second phase of Study 1 involved rating reminder delivery formats by their relative acceptability across various contextual scenarios, which sampled the space of high and low notification urgency, event importance, and user attention. It showed, first, that acceptability ratings for reminder formats were more positive when attention requirements were low and notification utility was high. It also showed that there was significant individual variability in the specific types of reminder formats preferred. Finally, it showed that often people expressed preferences for the types of reminder formats that had been consistently viewed as annoying in the first phase of the study. These preliminary findings encouraged us to further examine how preferences vary for different notification formats delivered in different contexts.

## 4 User Study 2: Multi-format Notifications in a Multi-task Environment

We conducted an experimental user study to explore the relative desirability of a set of three notification delivery formats when issued in different contexts. The study involved two computer games. One game, which was always available to the user to play, was the Memory game adapted from Gluck, *et al.* [7]; the other was called the “Alien” game and was an adaptation of the popular arcade game “Space Invaders” [13].<sup>1</sup> For ease of exposition, we adopt the terminology of Trafton, *et al.* [14] and refer to the Memory game as the “primary task” and the Alien game as “secondary.”

There were two sessions of the primary task, during which notifications were issued to alert participants to the availability of the secondary task. Upon receipt of a notification, participants were at liberty to switch to the secondary task or maintain focus on the primary task. Beforehand, we conducted an initial phase of the study that was used to inform our design decisions regarding the primary task load.

### 4.1 Differentiating Task Load

For this first phase of the study, we needed to establish whether there is a significant difference in the workload requirements of various configurations of game boards used in the Memory task. Each of ten participants played four versions of Memory. In two versions, the board was small (6x6), and the other two versions comprised larger boards of size 8x8. Within each size category, we varied the similarity of the images to be matched: they were either very distinct (different colors, shapes, sizes) or similar (with a shared color scheme or image theme). The order of the boards was counterbalanced among participants, and an interview was conducted after each set of two Memory boards.

---

<sup>1</sup> In our version of Space Invaders, there were no obstacles or shots fired back by the aliens: participants merely targeted and eliminated 2D moving alien spacecraft.

To evaluate the relative task load requirements of the different boards, NASA Task Load Index (TLX) [8] ratings, also used in [1,7,11] among others, were analyzed between each pair of boards using a paired *t*-test with an alpha-adjustment to account for multiple evaluations per board. We found no recognizable difference, statistically, between the two smaller boards on mental demand, effort or frustration (with all  $p > 0.4$ ). However, all other board pairings led to significant differences in these workload ratings ( $p < 0.05$  in all cases). Qualitative findings supported these results.

## 4.2 Multi-format Notifications in a Multi-task Environment

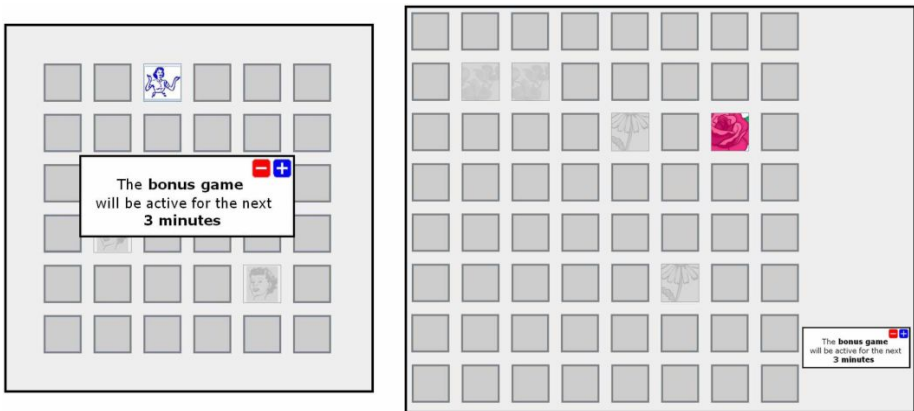
The purpose of our primary study was to measure the desirability of notifications delivered in different formats in a multi-task environment. As such, participants were asked to provide positive or negative feedback for each notification as it was delivered throughout the study.

**Participants.** We recruited 28 participants (9 female) between the ages of 18 and 49, all of whom reported that they use a computer for over an hour a day. Upon recruitment, participants were informed that they would receive compensation of \$5 per hour of participation and a \$10 bonus if they were more successful than two-thirds of participants at playing a set of computerized games.

**Tasks.** Similar to the first phase of our study described above, the objective of the Memory game was to find all pairs of matching images on each game board. Every match was worth 10 points toward the user's total score. Informed by findings from the initial phase of this study above, we divided Memory game boards into two levels of task load: one, the low-focus level, included twenty size 6x6 boards hiding 9 images (such that there were four identical instances of each image hidden on any given board); and another, the high-focus level, included five size 8x8 boards each hiding 32 image pairs that were much more similar to one another, with each matching pair differing only by a few minor visual characteristics from every other matching pair. In the Alien game, the user's objective was to remove as many aliens as possible from the screen during the time the game was active. There were two versions of the Alien game: the basic Alien game, in which all aliens were white in color and worth one point each, and the Bonus game, in which a handful of aliens were colored blue or red. Participants were informed that blue aliens were worth 20 points each, and that red aliens were worth 100 points. We designed these two games in such a way as to model two levels of importance, and with a variable amount of time for which a game is available, they represent two levels of urgency as well.

**Notifications.** Notifications were used to alert a user to a secondary task becoming available. A notification named the newly-available secondary task (either the Alien or Bonus game) and the exact amount of time for which it was to be available (either thirty seconds or three minutes). This was our method of representing the contextual features related to a notification: importance is high when the Bonus game is available (because it is worth far more points than can be attained in the Alien game) and low when the notification specifies the availability of the Alien game; and urgency is high when the game is available only for the next 30 seconds, whereas it is considered low if the user need not attend to the secondary task immediately.

Notifications were issued in one of three formats: two visual and one auditory. Our choice of visual notifications was informed by the two categories of reminder delivery formats from our initial experiment, Study 1. One visual notification was a large, occluding window that popped up in the center of the screen in which the user was playing the Memory game. The other visual notification was smaller and appeared toward the bottom right-hand corner of the screen. The design of these notification formats was strongly influenced by our initial experiment (Study 1) in which notification formats were grouped into two distinct equivalence classes (based on their relative levels of perceived intrusiveness). A third, auditory notification was a woman's voice dictating the same information that is presented in the visual notifications. We included this notification format to examine an alternative modality for information delivery. Fig. 1 depicts a low-focus board and an occluding notification on the left, and a high-focus board and a peripheral notification on the right.



**Fig. 1.** A small Memory board displaying an occluding notification (left), and a large Memory board with a peripheral notification (right)

Upon receipt of a notification, it was requested that participants rate the notification using a pair of positive and negative feedback buttons. A positive rating indicated a desirable notification format in the given context and content of the notification, and vice versa. To capture the positive and negative feedback that was requested of the participants upon receipt of each individual notification, the feedback buttons (depicted in Fig. 2) were placed to the left of the keyboard: a blue smiling face was used for positive feedback, and a red frowning face was used as negative feedback.

**Experimental Design.** This study comprised a 2 Task load (high- and low-focus Memory task) X 2 Importance (high, low) X 2 Urgency (high, low) X 3 Notification format (occluding, peripheral, auditory) mixed design. All of the treatments were within subjects, so that user preference information could be compared across all users, scenarios and notification types. We used a Williams design [5] to counterbalance all of the treatment combinations.

### 4.3 Protocol and Evaluation Metrics

**Procedure.** Participants were first presented with a consent form, a pre-study demographics questionnaire, and verbal and written instructions for participating in the study. Then they were familiarized with each of the notification formats and administered a practice session in which they were walked through the process of playing the Memory game, receiving a notification, providing feedback, and addressing the secondary task (the Alien or Bonus game). Participants were encouraged to ask questions throughout this orientation process.



**Fig. 2.** The feedback buttons used for rating each delivery format upon receipt of a notification

Each user then participated in two 30-minute task sessions. One session involved the low-focus Memory task, and the other the high-focus task. During each session, each of twelve possible notifications was issued. After providing feedback to a notification using one of the feedback buttons in Fig. 2, a virtual button would appear in the upper right corner of the screen that, upon being clicked, would start the secondary game.

In addition to the workload assessment, a structured interview was conducted after each session of the study. Participants were asked to describe the desirability of the three notification delivery formats, and to rank order each of them to the degree possible. Participants were also asked whether certain notifications seemed generally more helpful than others, and whether there were certain times during the session in which notifications were more useful. We were also interested in participants' perceptions of their attentiveness to notifications and whether there were occasions on which they chose to defer attendance to a secondary task or ignore it altogether.

**Subjective Measures<sup>2</sup>.** Along with positive and negative (reactive) feedback ratings for each notification, our subjective measures also include users' responses to the NASA Task Load Index survey, which was completed in the first phase of our study as well as after each session of our experiment, both for the primary and secondary tasks. We also conducted an interview after each session of the experiment, asking participants about their perceptions of notifications delivered during the session, both

---

<sup>2</sup> Objective measures were taken for the sake of comparison with prior work, but accounting for space, they will not be reported here.

alone and in combination with contextual features of the multi-task environment in which the study was conducted.

#### 4.4 Hypotheses

Our central hypotheses postulated that having access to a user's task context and the content of a notification will be enough to predict user response to the delivery format of that notification, and that users can accurately articulate their notification preferences both with and in the absence of contextual information.

**H1:** Contextual features (i.e., relative task load and notification utility information) will be sufficient to determine the desirability of each notification delivery format.

**H2:** Reactive ratings will correspond to reflective ratings; people who indicate an overall preference for a certain delivery format will provide more positive ratings to that format than to others. Furthermore, a delivery format that is favored by users overall (reflectively) will be provided the most positive ratings in a reactive setting.

## 5 Results

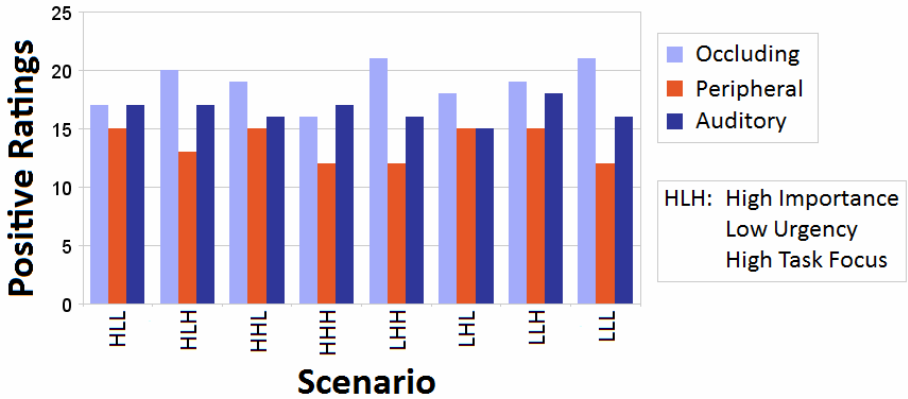
To address our first hypothesis, we analyze participants' reactive ratings in each scenario. Findings are displayed in Fig. 3, which depicts the number of positive ratings (on the y-axis) for the three notification formats in each of the eight contextual scenarios (the x-axis).

It can be noted that, contrary to the findings in the prior study, there is no strict trend in relative desirability from the low-load, high-utility scenario toward the high-load, low-utility scenario. The total number of positive ratings over all three notification formats was fairly stable over all eight scenarios, and it was even more stable when considering only visual notifications (to compare with Study 1): there was an approximately equal number of positive ratings over both occluding and peripheral notifications in every scenario. It is also the case that occluding notifications were almost always preferred to the other notification delivery formats, with peripheral notifications least positively rated in all scenarios. On the surface, this suggests that there is no support for hypothesis **H1**: the contextual features we examine do not seem to influence notification preferences, because occluding notifications are most preferred independent of contextual information.

Most importantly, we also found that there is again high individuality of preference ratings. Here, with only one type of notification style per category (highly intrusive, moderately intrusive and auditory), preference variation is exhibited between participants over the three notification categories. In particular, despite occluding notifications receiving the most positive ratings across scenarios (see Fig. 3), it is not the same set of users that provided those positive ratings in each scenario. Further, in the reflective setting, there was little overlap in the set of participants preferring, e.g., occluding notifications, across the high and low focus task sessions. Finally, for a majority of users, context does in a variety of ways affect their reactive ratings: only two participants in this study had static preferences across scenarios (and in both cases every notification was positively rated). It is thus *not* the case that there is any sort of uniform function from contextual features to preferences across users.



## Reactive Preferences



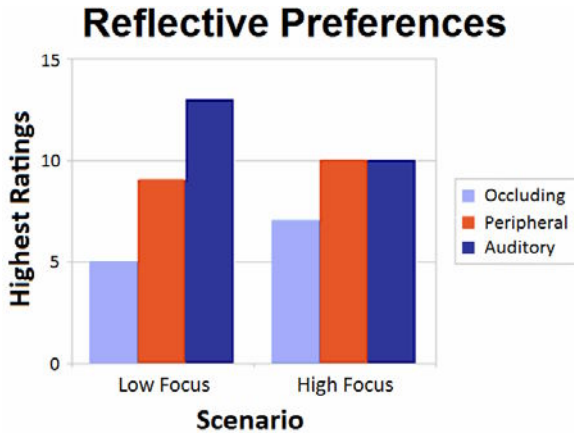
**Fig. 3.** The number of positive reactive ratings for each type of notification delivered in each of the eight contextual scenarios: {high, low} task load x {high, low} importance x {high, low} urgency. Each notification was rated individually; with 28 participants, any notification in any context could receive up to 28 positive ratings.

To address hypothesis **H2**, we compare reflective user ratings to the reactive ratings in Fig. 3, first on an aggregate basis and then individually. Aggregated reflective ratings are presented in Fig. 4, where ratings are grouped by partial orderings of the three notification delivery formats, as they were rank-ordered by participants during post-session interviews. With twenty-eight participants, we expected to accumulate 56 ranked orderings because each participant was interviewed twice: once after each of the two study sessions; however in two cases preference rankings could not be easily articulated,<sup>3</sup> which left us with 54 total rankings. With this in mind, from here forward, data will report the equivalent of twenty-seven participants rather than twenty-eight. The data presented in the graph shows the number of times a particular notification format was indicated to be preferred to the others after each session of the experiment.

When we categorize user preference ratings with respect to which type of notification was considered best, or most preferred, and which notification format was considered worst, or least preferred, overall, a resounding half of the user ratings placed occluding notifications last in their preferred list of notification formats. Further, Fig. 4 shows pictorially that occluding notifications were seldom preferred to both other types of notifications used in the study.

Individual user preferences—both reactive and reflective—are displayed in Fig. 5. The figure shows two rectangular diagrams, one displaying preference information elicited during the high-focus task session (upper diagram), and the other displaying preferences from the low-focus task session (lower diagram). Each column in the diagrams describes one individual user: the first row of both diagrams depicts that user's reflective preference for that task session, and the four boxes below a user's reflective preference describe that user's reactive ratings from each of the four contexts associated with that task session (all combinations of high and low importance and urgency). Users

<sup>3</sup> Two study participants believed their preferences to be directly linked with context and thus could not establish an overall ranking in one of the two sessions each.



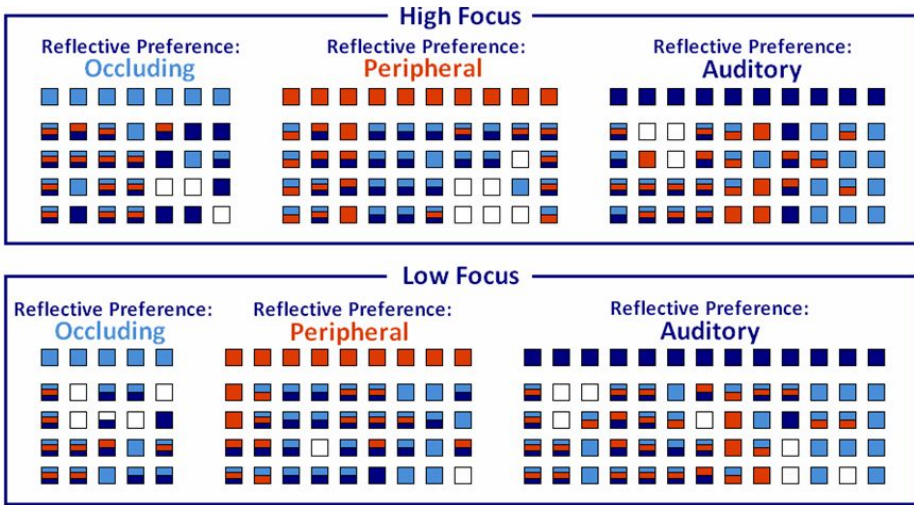
**Fig. 4.** The number of times Ocluding, Peripheral and Auditory notifications were ranked highest in participants’ reflective ratings, in all contexts with the low focus task (left) and high focus task (right)

are grouped by their reflective ratings: there are three groups, representing a reflective preference for ocluding (in light blue, and corresponding to the first row of boxes), peripheral (in orange, of a middle shade), and auditory (colored dark blue). As an example of reading the top diagram (the high focus session), it indicates by its last column of ratings that one user reflectively preferred auditory notifications to ocluding and peripheral during the post-session interview, but during that particular task session the user only provided positive ratings to ocluding notifications.

As an aid in understanding the data depicted in the diagram, we will define three preference conditions: *complete match*, *partial match*, and *total mismatch*. A user whose data falls into the complete match category has provided positive feedback in every scenario of the reactive setting to that notification format deemed most preferred in the reflective setting. This does not preclude other formats having been provided positive ratings as well – e.g., someone who prefers auditory to visual notifications may appreciate any type of notification format under certain conditions. In Fig. 5 above, the first participant (column) of each category of reflective preferences, both in the high and low focus scenarios, is representative of a complete match: in all six of these cases, that user consistently provided a positive rating to the notification for which he or she indicated highest overall preference.

On the other end of the spectrum, a total mismatch describes a participant who indicated an overall preference for a notification format that did not receive a single positive rating in any of the scenarios within a given session. One example from Fig. 5 is the rightmost column in both sessions, describing a participant who indicated an overall preference for auditory notifications but only provided positive ratings to ocluding notifications.

Finally, a partial match describes those participants who provided some but not a complete set of positive rating(s) to the notification format for which they indicated highest preference. This encapsulates all participants whose ratings do not fall into either the complete match or total mismatch categories (and it thus comprises the set complement of the union of the complete match and total mismatch sets).



**Fig. 5.** Individual reflective preferences paired with reactive preference ratings, grouped by task session (high and low focus): each column represents a single user; the top row of boxes in each session displays individual users' format preferences during that session (corresponding to the groupings indicated by the reflectively preferred notification format), and the four boxes beneath each user's reflective preference present a visual description of that user's reactive preference ratings: light blue signifies a positive rating for the ocluding notification delivered in the given context, orange indicates a positive rating for the peripheral notification, and dark blue indicates a positive rating for the auditory notification. If more than one notification was rated positively, multiple colors share the space. Empty spaces (i.e., white boxes) indicate that no notification was rated positively in that context.

Upon first glance at the pair of diagrams in Fig. 5, there is very little correlation between reflective and reactive user ratings. There are twenty-two complete match cases, in which people's reactive preferences correspond to their reflective preferences. Yet, this does not even account for half of the 54 sets of preferences that were acquired, suggesting that there is only a weak, if any, basis for supporting hypothesis **H2**. More notably, in fact, is the number of people whose preferences differ between reactive and reflective ratings. The number of participants who fall into the total mismatch category across both sessions is seventeen, almost as many as there are complete matches; and in total there are thirty cases (15 in each session, accounting for more than half of the participants) in which the reflectively-preferred notification format was not the one to receive the greatest number of positive ratings in the reactive stage (this includes fifteen cases of the partial match condition).

## 6 Discussion

Our results imply that the desirability of a notification delivery format cannot be generalized across either computer users or contextual settings. While feedback ratings provided directly in response to each notification indicated that ocluding notifications would satisfy more users than either of the other two delivery formats, more

participants rated the peripheral and auditory notifications most desirable overall (i.e., across contexts). And a number of participants provided contradictory preference information between the reactive and reflective settings. This suggests that user preferences may not be easily articulated, at least in the absence of contextual grounding.

Results also suggest that designers wishing to embed user preferences into intelligent software tools could potentially elicit that information by asking users to provide reactive feedback to explicit contextual scenarios (similar to the second phase of Study 1), but exhaustively enumerating all possible contexts would be a cumbersome if not intractable undertaking. Furthermore, because context does not appear to influence user preferences in a similar way across the majority of users, information acquired from one user would not necessarily generalize to any other user, meaning that the process would have to be replicated for each individual user. Instead, we propose that machine learning and user modelling techniques can perform this preference learning automatically, by learning the contextual features most relevant in understanding the effects of context and notification format on individual user preference patterns and then customizing notification delivery accordingly.

One important limitation of our primary study is that it was conducted by way of a pair of computer games. A gaming environment is not necessarily representative of an actual workplace, toward which much of our work is targeted. While our long-term goal is to understand notification preferences in real-world settings, we used a game-playing scenario as an initial test of our hypotheses within a controlled experimental setting. Further studies are currently underway in a technical support center, where actual notifications are being judged by actual employees toward our goal of understanding the similarities and differences among individual user preferences for notification delivery.

Unlike [9], I did not control the timing of a notification by matching task interruptions to known task breakpoints or otherwise. However, a number of participants commented (independently of the interview question regarding timing) on timing as a factor influencing their reactions to notifications. Interestingly, however, these comments pointed in different directions, with some users indicating a preference to receive notifications at the beginning or end of the primary task and others preferring notifications to arrive in the middle, with specific comments relating to effects of the games on that user's emotional state. This indicates that the tasks in which a user is engaged or has the opportunity to be engaged may themselves be important factors influencing user preferences for different notification delivery formats.

## 7 Conclusion and Future Work

With the surge in technology use over the past decade, people's attention is becoming more scarce; we are more often interrupted from our daily activities, both computer-based and not, and thus potentially inured to the effects of certain notifications, essentially making us less susceptible to them causing a disturbance in our daily routine. As a result, it is likely that people may be affected by notifications in a variety of different ways. Our results support this claim by suggesting that different people react and respond in different ways to different types of computerized notifications. Using only the contextual task and notification information selected for consideration in this

study, our experiments with a non-adaptive, non-personalized system led to the satisfaction of between half and three-quarters of users, by way of issuing notifications in an intrusive manner across all contextual settings. User satisfaction levels can likely be improved, supporting the need for user modelling and/or machine learning tools, e.g., similar to those used in [6, 7] that have the ability to learn desirability of notification formats directly from user response and feedback.

The findings from our experiment motivate future study into the desirability and performance effects of a broader selection of notifications, from computer-based to mobile platforms, taking into consideration a broader selection of contextual features. Considering notification utility and a single aspect of task context (attentional focus requirements of the current task) is just one combination of contextual features that may influence a user's notification preferences. An exploration of other features, in a wider range of granularities, will be useful for understanding whether other aspects of context are better predictors of user preferences.

**Acknowledgments.** The authors would like to especially thank Jennifer Gluck for her kind support of this project, both in discussions and for materials that she donated for the sake of standardization and comparison. Thanks are also due to Dr. Joanna McGrenere for her early assistance with the statistical analysis of our results. And a special thanks to our blind reviewers for their insightful comments and suggestions, as well as our study participants for the time and energy they each devoted to the project.

## References

1. Adamczyk, P.D., Bailey, B.P.: If not now, when?: the effects of interruption at different moments within task execution. In: CHI 2004: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 271–278. ACM, New York (2004)
2. Cutrell, E., Czerwinski, M., Horvitz, E.: Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. In: Interact (2001)
3. Czerwinski, M., Cutrell, E., Horvitz, E.: Instant messaging: Effects of relevance and time. In: People and Computers XIV: Proceedings of HCI 2000, vol. 2, pp. 71–76. British Computer Society (2000)
4. Edwards, A.: Balanced latin-square designs in psychological research. *American Journal of Psychology*, 598–603 (1951)
5. Fogarty, J., Hudson, S.E., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C., Yang, J.: Predicting human interruptibility with sensors. *ACM Trans. Comput.-Hum. Interact.* 12(1), 119–146 (2005)
6. Gajos, K.: Automatically generating personalized user interfaces, phd dissertation (2008)
7. Gluck, J., Bunt, A., McGrenere, J.: Matching attentional draw with utility in interruption. In: Proceedings of CHI 2007, pp. 41–50 (2007)
8. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P., Meshkati, N. (eds.), Elsevier Science, Amsterdam (1988)
9. Iqbal, S.T., Bailey, B.P.: Effects of intelligent notification management on users and their tasks. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2008), Florence, Italy (2008)

10. Levitt, J.: Internet zone: Good help is hard to find. Information Week: Listening Post (2001), <http://www.informationweek.com/835/35uwjl.htm>
11. Mark, G., Gudith, D., Klocke, U.: The cost of interrupted work: more speed and stress. In: CHI 2008: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, pp. 107–110. ACM, New York (2008)
12. McCrickard, D.S., Chewar, C.M.: Attuning notification design to user goals and attention costs. *Communications of the ACM* 46, 67–72 (2003)
13. Nishikado, T.: *Space invaders* (1978)
14. Trafton, J.G., Altmann, E.M., Brock, D.P., Mintz, F.E.: Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *Int. J. Hum.-Comput. Stud.* 58(5), 583–603 (2003)
15. Vastenburg, M.H., Keyson, D.V., de Ridder, H.: Considerate home notification systems: A field study of acceptability of notifications in the home. *Personal and Ubiquitous Computing* (2007)
16. Weber, J.S., Pollack, M.E.: Evaluating user preferences for adaptive reminding. In: CHI Extended Abstracts (2008)